Proc. Nat. Acad. Sci. USA
Vol. 71, No. 2, pp. 283-287, February 1974

A Proposed Model for Interaction of Polypeptides with RNA

(precellular evolution/L-amino acids in proteins/double-helical conformations/twisted-β-sheets)

CHARLES W. CARTER, JR.* AND JOSEPH KRAUT

Department of Chemistry, University of California, San Diego, La Jolla, Calif. 92037

Communicated by Stanley L. Miller, September 21, 1973

ABSTRACT **Pairs of antiparallel β polypeptide-chain segments in known protein structures are usually observed to form right-handed double helixes with helix parameters in the same range as those of nucleic acids. We have constructed a model containing only standard bond lengths, bond angles, and dihedral angles in which such a polypeptide double helix fits precisely into the minor groove of an RNA double helix with identical helix parameters. The geometry of the RNA portion is essentially a hybrid between those of the A and A′ forms. Hydrogen bonds can be made between the ribose 2′-hydroxyls and polypeptide carbonyl oxygens. Since such precise complementarity between the stable conformations of RNA and polypeptides is unlikely to be merely coincidental, we propose that it played a fundamental role in the initiation of precellular evolution. Specificially, we propose that the two double-helical structures are mutually catalytic for assembly of one another from activated precursors in the prebiotic soup, and moreover that they provide some degree of genetic coding.**

There is probably a fundamental structural complementarity between RNA and polypeptides; this complementarity must have played a crucial role in the earliest stages of biological evolution. A recent observation concerning the conformation of β-sheet structures in proteins suggested to us a possible basis for such complementarity. It now is evident that the stable conformation of an extended polypeptide chain is somewhat twisted in a right-handed sense (1). Thus, a pair of antiparallel β chains will form a "polypeptide double helix" with a relatively long pitch. Such structures, with pitch and radius *roughly* in the same range as those of nucleic-acid double helixes, are quite common in proteins. Inspection of stereo pictures drawn from atomic coordinates revealed that they occur in chymotrypsinogen (2) at residues 31–45, 44–55, 197–214, and 211–230; in subtilisin (3) at residues 202–219; in HiPIP (4) at residues 48–64 and 59–73; in pancreatic trypsin inhibitor (5) at residues 18–35; in cytochrome b_5 (6) at residues 21–32; in ribonuclease S (7) at residues 78–87 paired with residues 96–105; in lactate dehydrogenase (8) at residues 266–275 paired with residues 286–295; in lysozyme (9) at residues 42–54 and 51–61; and in carboxypeptidase (10) at residues 32–53. These observations prompted us to search, by means of careful model building, for a polypeptide double helix that would be *precisely* complementary to double-stranded RNA. This communication describes such a structure, and its implications for a possible origin of biological evolution.

Model building

We first constructed a Kendrew–Watson skeletal model of the A form of RNA using published atomic coordinates (11). This model consisted of a double-helical segment containing seven base pairs, which we felt was sufficiently long to reveal obvious discrepancies between the respective helix parameters of the polynucleotide and any potential complementary polypeptide structures. Alternatively we might have started with several other related RNA structures, specifically the A′ form of RNA (11) and helixes derived from the dinucleotide structures GpC (12) and ApU (13). These structures exhibit helix parameters ranging from 10 to 12 residues per turn and from 26 Å to 36 Å in pitch. We started with the A form simply because it seemed most likely to represent a native structure (11). However, we anticipated that some adjustment would be required, and the existence of the other forms gives an indication of the permissible range of such adjustments.

Next we considered what fundamental limitations might be imposed upon a polypeptide helix by the requirement that it be complementary to this RNA structure. Clearly it must be a right-handed helix. Moreover, in order to maintain the 2-fold axes, it is necessary that the polypeptide helix also have 2-fold axes, and that they coincide with those of the polynucleotide. We are thus immediately limited to consideration of right handed double-stranded antiparallel β structures of precisely the type commonly observed in proteins.

Into which of the two grooves on the RNA might this polypeptide double helix best fit? Although the "major" and "minor" grooves in RNA are about the same width, an important distinction is that the backbone phosphate groups project into the major groove, whereas the ribose rings project into the minor groove. Consequently the major groove is lined with phosphate oxygen atoms as potential hydrogen-bond acceptors, while the minor groove offers projecting 2′-hydroxyl groups as both potential hydrogen-bond donors and acceptors. However, the phosphate groups are so positioned that it seemed difficult to insert a polypeptide while maintaining a repeating hydrogen bond from the backbone amido NH group. On the other hand, the 2′-hydroxyl groups seemed to be spaced at about the right distance to donate hydrogen bonds to the backbone carbonyl oxygen atoms. For this reason we decided to explore in detail only the polypeptide-binding possibilities of the minor groove.

There are only two fundamentally different ways in which a polypeptide double helix can be fitted into the minor groove. Specifically, adjacent polypeptide and polynucleotide chains may be either parallel or antiparallel. In either case an "asymmetric unit" consists of one nucleotide ribose phosphate and a dipeptide backbone segment. Also, in either case one amino-acid side chain of the dipeptide segment will point in toward the helix axis, the other will point out, and the dihedral angles ϕ and ψ of the two amino acids will be slightly different. The most important distinction between these two ways of inserting the polypeptide chains into the minor groove is that, in

* Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge, England.

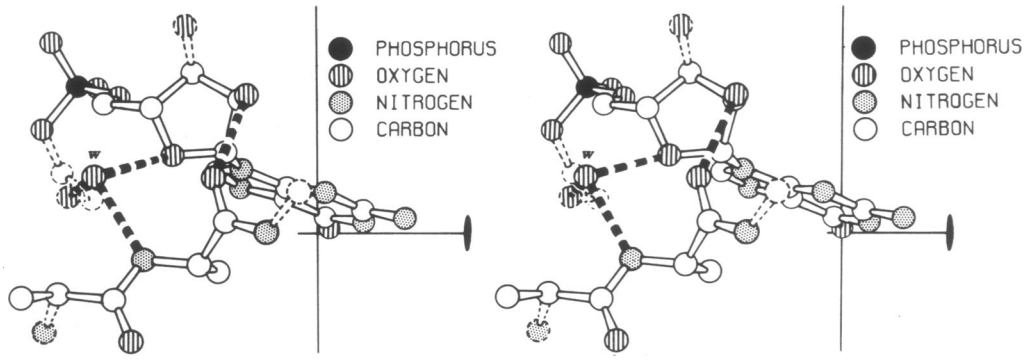

FIG. 1. An asymmetric unit of the proposed complex. Atoms that precede and follow this asymmetric unit are indicated by *broken lines*. Hydrogen bonds are indicated by *heavy dashed lines*. We have also included in this figure a water molecule (*W*) which can form a hydrogen-bonded bridge between N2 in the polypeptide backbone; O5, the furanose ring oxygen; and O6, the 2′-hydroxyl oxygen from the ribose of the preceding asymmetric unit. *Vertical* and *horizontal lines* indicate the helix axis and a 2-fold axis, respectively.

one case, each inward-pointing pair of β-carbon atoms opposite one another on antiparallel strands (see Fig. 2B and C) will be separated by about 3.5 Å; in the other case they will be separated by about 4.5 Å. The second choice results in unacceptably close van der Waals contacts which we were unable to alleviate by making such small adjustments as are possible while maintaining coincidence of the polypeptide and polynucleotide 2-fold axes. Furthermore, the hydrogen-bonding possibilities between polypeptide and polynucleotide seemed less favorable with this choice. Therefore, we concentrated our attention on building a model based upon the first choice, that is, the one with adjacent polypeptide and polynucleotide chains parallel, and with the shorter distance between inward pointing β-carbon pairs.

One more binary choice must be made concerning the relative positioning of the polypeptide and polynucleotide chains. The RNA double helix contains a 2-fold axis at the origin of each asymmetric unit, and 2-fold axes half way between these. Similarly, the polypeptide double helix also contains two sets of 2-fold axes. Thus, there are two ways in which the respective sets of 2-fold axes can be made to coincide. With one choice, the available polypeptide backbone NH group is about 4 Å from the ribose 2′-hydroxyl oxygen atom, and therefore, too far away to form a hydrogen bond. With the second choice, the available polypeptide backbone carbonyl oxygen atom is about 2.5 Å from the ribose 2′-hydroxyl oxygen atom. It appeared that with slight further adjustments this arrangement could probably be made into a convincing structure,

TABLE 1. *Cylindrical polar coordinates for the atoms in an asymmetric unit*

	r(Å)	θ(°)	Z(Å)		r(Å)	θ(°)	Z(Å)
Ribose phosphate				Polypeptide			
OR6	10.30	−28.3	3.73	CA1	12.57	−50.8	−1.63
CR1	8.75	−34.4	2.25	CB1	14.10	−50.8	−1.70
CR2	8.98	−31.3	3.68	C1	12.08	−44.1	−1.78
CR3	9.02	−39.8	4.45	O1	12.16	−41.5	−2.89
CR4	9.85	−44.8	3.46	N2	11.57	−41.1	−0.71
OR5	9.36	−42.6	2.16	CA2	11.13	−34.1	−0.85
CR5	9.94	−53.5	3.61	CB2	9.62	−34.0	−1.06
OR1	9.72	−70.1	2.98	C2	11.56	−29.8	0.31
OR2	9.66	−67.7	5.46	O2	11.41	−31.6	1.48
OR3	7.62	−73.3	4.27	N1	12.11	−24.1	0.02
P1	8.83	−67.4	4.23				
OR4	8.65	−57.2	3.87	H_2O	11.75	−47.9	1.54
Purine				Pyrimidine			
N9	7.34	−34.4	1.80	N1	7.34	−34.4	1.80
C8	6.54	−43.7	1.76	C6	6.69	−44.4	1.84
N7	5.33	−42.4	1.31	C5	5.44	−46.8	1.42
C5	5.42	−27.9	1.00	(O4,N4)	3.52	−33.3	0.51
(N6,O6)	3.49	−13.1	0.13	C4	4.72	−33.8	0.92
C6	4.72	−15.5	0.46	N3	5.62	−22.5	0.90
N1	5.70	−4.9	0.30	O2	7.81	−18.1	1.28
C2	6.97	−7.7	0.64	C2	6.91	−24.6	1.32
N3	7.50	−16.4	1.14				
C4	6.73	−25.2	1.29				
(Gua,N2)	7.96	−0.4	0.41				

Helix-generating parameters: $\Delta\theta = 30.8°$; $\Delta Z = 2.71$ Å. Two-fold axes at $\theta = 0$, $Z = 0$ and $\theta = 15.4°$, $Z = 1.36$ Å.

with hydrogen bonds from the 2′-hydroxyls to the carbonyl oxygens.

After we had constructed the polypeptide backbone and fitted it to the RNA, several minor changes in both parts were necessary in order to insure that both RNA and polypeptide would have the same helix-generating parameters, $\Delta\theta$ and ΔZ. These changes were made with the aid of a digital coordinate measuring device (14). An intermediate set of atomic positions was then obtained by averaging cylindrical coordinates over the entire model containing fourteen asymmetric units. Finally, these coordinates were further adjusted by computerized least-squares procedures so as to conform with standard geometry. Standard purine, pyrimidine, and ribose rings and a standard phosphate group (15) were substituted into the RNA portion of the model. Standard polypeptide chain geometry (16) was incorporated directly with the aid of a program written by S. T. Freer and R. A. Alden.

Description of the model

Atomic coordinates for an asymmetric unit and helix generating parameters are given in Table 1. The asymmetric unit is illsutrated in Fig. 1. Deviations of covalent bond lengths and angles from the standard values given in refs. 15 and 16 are all less than 0.01 Å and 6°. Repeated application of computerized least-squares procedures would probably have been capable of producing atomic parameters arbitrarily close to any set of standard bond lengths and angles, but we felt nothing would be gained by further refinement. Similarly the RNA backbone dihedral angles are all within the range observed in nucleotides (17, 18), and the polypeptide dihedral angles $(\phi_1,\psi_1) = (-120°, 128°)$, $(\phi_2,\psi_2) = (-143°, 144°)$ differ by not more than 5° from those of the standard β_P and β conformations (16). The closest van der Waals contact is between O2 (the carbonyl oxygen of the second amino acid) and CR1 (C1 of the

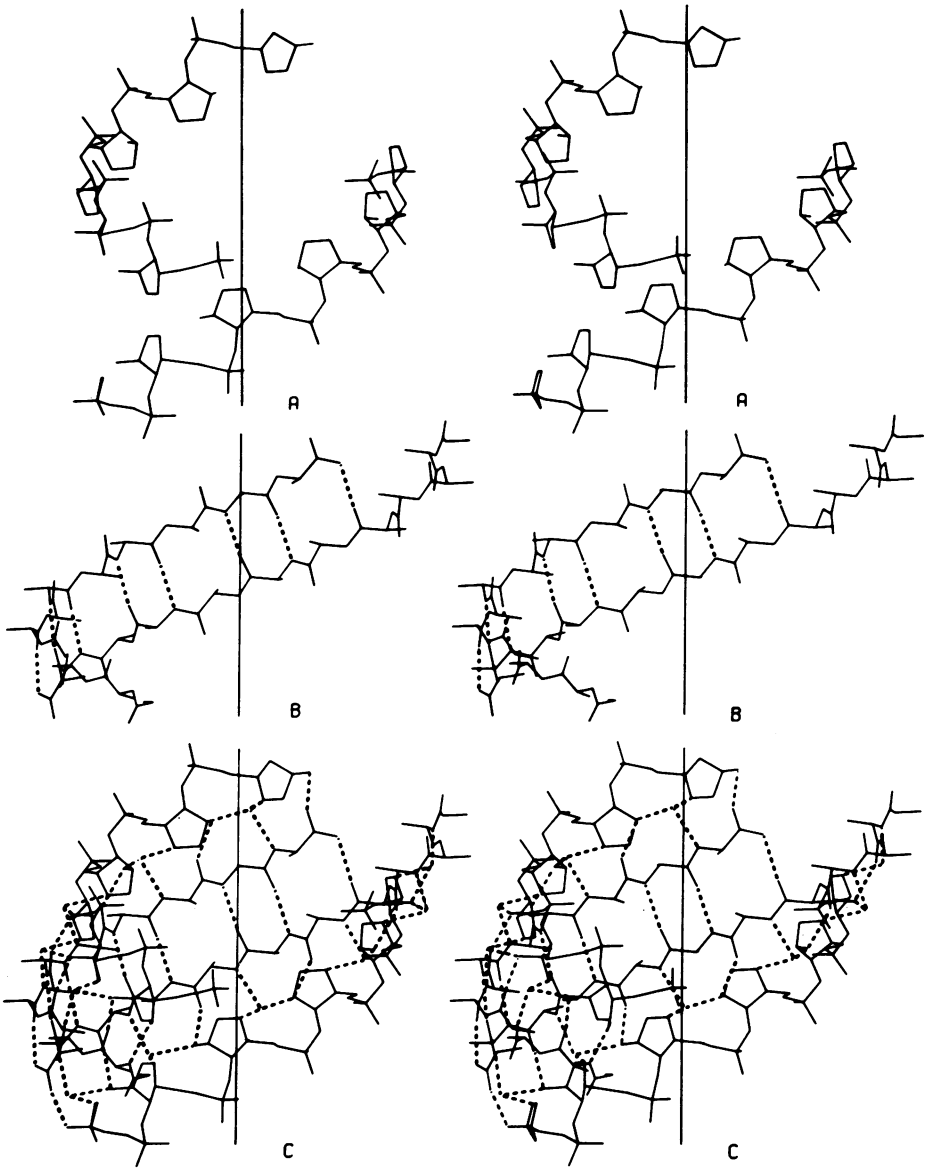

FIG. 2. Backbone portions of polynucleotide and polypeptide double helixes generated from the asymmetric unit. (*A*) The ribose phosphate backbone. Base pairs have been omitted for clarity. (*B*) The polypeptide backbone, including β-carbon atoms. (*C*) The complex formed between structures *A* and *B*.

TABLE 2. *Characteristic helix parameters of RNA*

	RNA-A	Fig. 2A	RNA-A′
Pitch	31 Å	32 Å	36 Å
Radius*	9.07 Å	9.26 Å	9.45 Å
Residues/turn	11	11.7	12
Base-plane tilt	15.0°	20.7°	8.9°

* Mean radius of ribose phosphate atoms.

ribose ring) which is 2.80 Å, or 0.1 Å *longer* than the shortest approach allowed by Ramachandran (16).

In Fig. 2*A* and *B* we have illustrated the backbone portions of polynucleotide and polypeptide double helixes generated from an asymmetric unit using the helix-generating parameters $\Delta\theta = 30.8°$ and $\Delta Z = 2.71$ Å and the 2-fold axes. The complex between them appears in Fig. 2*C*. Three remarkable features of this model complex provide compelling evidence that it could indeed represent a real structure:

(*1*) Owing to the fact that we have been able to introduce into the asymmetric unit bond lengths and bond angles that are arbitrarily close to their standard values, while simultaneously defining the helix-generating parameters $\Delta\theta$ and ΔZ to be identical for the ribose-phosphate and polypeptide backbone chains, the steric relationships for the asymmetric unit shown in Fig. 1 can be maintained along a helix of arbitrary length.

(*2*) The polynucleotide and polypeptide backbone chain conformations in our model lie very close to the accepted free-energy minima. Let us consider each one in turn. How does the RNA double helix shown in Fig. 2*A* compare with the known structures of RNA-A and RNA-A′ (11)? We may examine four characteristic parameters of each helix: its pitch, its radius, the number of residues per turn, and the base-plane tilt. Our RNA helix appears to be a hybrid between the A and A′ helixes (Table 2). In fact, two of the seven characteristic dihedral angles lie between their corresponding values in RNA-A and A′, while the remaining five lie 7° or less outside this range. Turning to consider the double-helical polypeptide portion of our proposed structure (Fig. 2*B*), we note that deviations of the ϕ and ψ angles of our model from those of the standard β and β_P conformations are evidently characteristic of the right-handed twist predicted by Chothia (1) to be most stable. Very similar structures are observed in many proteins; we have found examples in every protein structure we have examined except myoglobin and the *c*-type cytochromes.

(*3*) These two structures fit together with remarkable precision (Fig. 2*C*), forming a repeating hydrogen bond between the ribose 2′-hydroxyl group and one of the two polypeptide carbonyl oxygen atoms. The geometry of this hydrogen bond is within expected limits. Its length is 2.6 Å and the OR6 atom lies 1.4 Å below the peptide plane. Both of these features appear to be stereochemically tolerable (19); moreover we believe the model could be further adjusted to improve them. In addition, it is apparently possible to introduce a hydrogen-bonded water molecule bridging alternate backbone NH groups to 2′-hydroxyl groups and furanose ring oxygen atoms. Although such hydrogen bonds between water and furanose ring oxygen atoms are not common, at least one example exists in the crystal structure of α-D-2′amino-2′-deoxyadenosine monohydrate (20).

We conclude that there is a fundamental complementarity between polynucleotides and polypeptides, and that it results directly from the stereochemistry of the two classes of biopolymers. It seems unlikely to us that such complementarity is merely coincidental; rather we feel that it may be the basis for interaction between proteins and double-stranded RNA.

One important feature of this proposed complex remains to be described. When the model is constructed with only β-methyl groups as inward-pointing polypeptide side chains, there is an empty region centered at $\theta = 15.4°$, $Z = 1.36$ Å, and $r \cong 9.7$ Å, and at all equivalent positions along the helix. It is sufficiently large to accommodate one or perhaps two additional side-chain atoms. This region is surrounded by four β-carbon atoms and the edges of two successive base pairs. Thus it sees four possible combinations of pyrimidine carbonyl O2 atoms and purine ring N3 atoms. Evidently in any real structure these regions would tend to be occupied. It should be noted that whenever a purine is guanine, its N2 amino group prohibits the presence of at least one inward-pointing β-carbon atom. Within the limits imposed by this constraint various such arrangements can be envisioned that would provide more detailed complementarity between specific polynucleotide and polypeptide sequences and, hence, some degree of genetic coding. However, there is no evident relationship between this potential coding device and the present triplet code.

An hypothesis for the initiation of precellular evolution

The model-building experiment provides an intriguing clue to what may have been the earliest events of biological evolution.

The crux of our hypothesis can be stated as follows: in addition to being structurally complementary, these RNA and polypeptide double helixes are also both primordial "polymerases" for one another. That is the RNA double helix catalyzes condensation of polypeptide from some activated precursory monomer, and reciprocally, the polypeptide double helix catalyzes condensation of polynucleotide from some other precursor, or perhaps the same one, e.g., aminoacyl nucleotides (21). Furthermore, we envision the catalytic process as being a repeating cycle in which the RNA–polypeptide complex can dissociate and each one can subsequently promote assembly of additional copies of the other. The degree of catalysis by both members need only be sufficient to insure that more than one copy of each is made before spontaneous degradative reactions, e.g., hydrolysis, inactivate the parent molecules.

Our model implies there will be some degree of fidelity in the hypothetical "polymerase" activity of both components, but ample opportunity for making errors and hence "mutants." Inevitably some mutants will be better catalysts than others, thereby providing a selection mechanism. Thus, the evolutionary cycle of reproduction, variation, and selection could have operated in a rudimentary fashion once the first double-helical polypeptide and/or polynucleotide was assembled by random events in the prebiotic soup. Specifically, we are proposing that a stage of biological evolution, involving both RNA and polypeptides in free solution, could have preceded the advent of replication and translation as we know them today.

We can also imagine, in addition to the polymerase activity, mechanisms whereby the original molecules are elongated by adding monomers or by joining short segments. Such mecha-

nisms might operate whenever dissociated "polymerase" and "product" happened to rejoin out of phase—i.e., literally somewhat unscrewed—and/or inverted with respect to one another. In this way the primordial population of biopolymers would grow in length and complexity, as well as in numbers, eventually giving rise to *single-stranded* molecules with internally complementary regions and, hence, with tertiary structure. Tertiary structure is now recognized to be an important feature in tRNA (25) and viral RNA (26) as well as in proteins. In this manner, our hypothesis provides a reasonable route to later stages of biological evolution.

This hypothesis provides particularly satisfying answers to several questions concerning the origin of some obvious, but nevertheless fundamental, properties of biological polymers.

(*1*) Why are there only L-α-amino acids in proteins, whereas the "primordial soup" must have contained racemic D,L mixtures as well as β-amino acids, all of which are incorporated into thermally prepared polypeptides (22)? The simplest explanation would be that the polypeptide component of the first self-reproducing system had to be some sort of regular helix. We propose that this hypothetical helix is the one represented in Fig. 2*B* and that the mutual interaction between this polypeptide helix and the RNA helix remained crucial throughout subsequent evolution.

(*2*) A similar question might be asked about the nucleic acids. The primordial soup must have contained racemic D,L mixtures of ribose derivatives. In this case the constraints imposed by the RNA double helix have long been obvious. A more interesting problem arises from the observation that condensation experiments (23, 24)) with activated nucleotides yielded only small quantities of the naturally occurring 3′–5′ linkages, the principal products containing instead 2′–5′ and 5′–5′ linkages. Moreover this is true even when condensation of 5′-phosphoadenosine with adenosine is carried out in the presence of a poly(U) template (24). The problem is obviated, however, if our hypothesis is correct. We would propose that primitive biological polynucleotides must be constructed exclusively with 3′–5′ linkages because only then will these polynucleotide helixes be structurally complementary to suitable polypeptides. Thus only such polynucleotide helixes will be able to act as polypeptide polymerases, and conversely, it is this linkage that will be synthesized on a suitable polypeptide template.

(*3*) Why is the genetic material nucleic acid and why are proteins the principal gene product through which selection acts? Our model suggests that this relationship would have characterized biological evolution even in its earliest stages. Although by our hypothesis both polypeptides and polynucleotides could serve as primitive polymerases for one another, there is a distinct difference in the fitness of these two polymers to carry a genetic message. Specifically a polynucleotide with *any* sequence of base pairs can be fitted into the model, whereas this is true only for a very restricted class of polypeptides. Indeed, only polypeptides in which every second side chain is small can fit into the RNA groove. This implies that *any* RNA sequence can code for some polypeptide product, but only certain polypeptides can act as RNA polymerases. Consequently, variation will occur most efficiently through the RNA sequence because all possibilities are translatable, whereas many variations in the polypeptide sequence will not be functional.

We conclude with a brief note concerning the relationship of DNA to our model and the above hypothesis. Since DNA lacks a 2′-hydroxyl group, it is unlikely that a polypeptide double helix will bind in the minor groove of the A form. In the B form of DNA the minor groove is narrower and therefore even less likely to provide a polypeptide-binding site of the type proposed here, but binding in the major groove cannot be excluded.

This work was supported by Research Grants GM 10928 from the National Institutes of Health and GB-30828X from the National Science Foundation. We thank Jon Peacock for capable assistance; Stephan Freer for his programming expertise generously given; Edward Stonebraker for measuring coordinates; and Mary Crawford for help with the illustrations.

1. Chothia, C. (1973) *J. Mol. Biol.* **75**, 295-302.
2. Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T. & Xuong, Ng. H. (1970) *Biochemistry* **9**, 1997-2009.
3. Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D. & Wright, C. S. (1971) *Biochim. Biophys. Res. Commun.* **45**, 337-343.
4. Carter, C. W., Jr., Kraut, J., Freer, S. T., Xuong, Ng. H., Alden, R. A. & Bartsch, R. G. (1974) *J. Biol. Chem.*, in press.
5. Huber, R., Dietmar, K., Rühlmann, A. & Steigemann, W. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 141-150 (atomic coordinates by personal communication).
6. Mathews, F. S., Argos, P. & Levine, M. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 387-395 (atomic coordinates by personal communication).
7. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245**, 305-328.
8. Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Lentz, P. J., Jr., McPherson, A., Jr., Schevitz, R. W. & Smiley, I. E. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 179-191 (atomic coordinates by personal communication).
9. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T. Phillips, D. C. & Sarma, V. R. (1965) *Nature* **206**, 757-761 (atomic coordinates by personal communication).
10. Quiocho, F. A. & Lipscomb, W. N. (1971) *Advan. Protein Chem.* **25**, 1-78.
11. Arnott, S., Hukins, D. W. L. & Dover, S. D. (1972) *Biochem. Biophys. Res. Commun.* **48**, 1392-1399.
12. Day, R. O., Seeman, N. C., Rosenberg, J. M. & Rich, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 849-853.
13. Rosenberg, J. M., Seeman, N. C., Kim, J. J. P., Suddath, F. L., Nicholas, H. B. & Rich, A. (1973) *Nature* **243**, 150-154.
14. Salemme, F. R. & Fehr, D. G. (1972) *J. Mol. Biol.* **70**, 697-700.
15. Arnott, S., Dover, S. D. & Wonacott, A. J. (1969) *Acta Crystallogr. Sect. B* **25**, 2192-2206.
16. Ramachandran, G. N. & Sasisekharan, V. (1968) *Advan. Protein Chem.* **23**, 283-437.
17. Sundaralingam, M. (1969) *Biopolymers* **7**, 821-860.
18. Arnott, S. & Hukins, D. W. L. (1969) *Nature* **224**, 886-888.
19. Donohue, J. (1968) in *Structural Chemistry and Molecular Biology*, eds. Rich, A. & Davidson, N. (Freeman, San Francisco), pp. 443-465.
20. Rohrer, D. C. & Sundaralingam, M. (1970) *J. Amer. Chem. Soc.* **92**, 4956-4962.
21. Katchalsky, A. (1973) *Naturewissenschaften* **60**, 215-220.
22. Saunders, M. A. & Rohlfing, D. L. (1972) *Science* **174**, 172-173.
23. Schramm, G. (1965) in *The Origins of Prebiological Systems*, ed. Fox, S. W. (Academic Press), pp. 299-315.
24. Sulston, J., Lohrmann, R., Orgel, L. E. & Miles, H. T. (1968) *Proc. Nat. Acad. Sci. USA* **59**, 726-733.
25. Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzerl, J. & Rich, A. (1973) *Science* **179**, 285-288.
26. Mills, D. R., Kramer, F. R. & Spiegelman, S. (1973) *Science* **180**, 916-927.